

3/9/1
DIALOG(R)File 347:JAPIO
(c) 2005 JPO & JAPIO. All rts. reserv.

02312464 **Image available**
DOCUMENT PROCESSOR

PUB. NO.: 62-229364 [JP 62229364 A]
PUBLISHED: October 08, 1987 (19871008)

INVENTOR(s): IWAI ISAMU

DOI MIWAKO
OKAMOTO TOSHIO

APPLICANT(s): TOSHIBA CORP [000307] (A Japanese Company or Corporation), JP
(Japan)

APPL. NO.: 61-072161 [JP 8672161]

FILED: March 29, 1986 (19860329)

INTL CLASS: [4] G06F-015/20

JAPIO CLASS: 45.4 (INFORMATION PROCESSING -- Computer Applications)

JAPIO KEYWORD: R139 (INFORMATION PROCESSING -- Word Processors)

JOURNAL: Section: P, Section No. 681, Vol. 12, No. 95, Pg. 157, March
29, 1988 (19880329)

ABSTRACT

PURPOSE: To attain the automatic evolution and output of the document data in the form of a specified layout and according to a layout rule corresponding to the logic structure, by obtaining the logic structures of the chapters, paragraphs, itemization, the ends of paragraphs, etc., from the data of inputted sentences.

CONSTITUTION: The document data supplied from an input part is stored in an original memory part 3 via a document control part 1. At the same time, a segmenting code, e.g., a renewing row code is detected out of the document data and a group of sentences which are segmented by the segmenting codes are successively extracted. Here the length of a single sentence is also measured. A header extracting part 6 decides whether or not an extracted sentence is possibly used as a header based on the length information of the single sentence and with reference to a headword dictionary 6a. Thus a header candidate is obtained if said sentence complies with a standard. This header candidate is compared with the header rule stored in a header rule dictionary 7a through a header deciding part 7. Then a header sentence is decided if the header candidate is coincident with the header rule. While a document text, like the end of a paragraph, etc., is decided if no header sentence is confirmed.

(19)日本国特許庁(JP)

(12)公開特許公報(A)

(11)特許出願公開番号

特開昭62-229364

(43)公開日 昭和62年(1987)10月8日

(51) Int. C l. ⁵	識別記号	府内整理番号	F I	技術表示箇所
G 06 F 15/20	3 0 1	H		
G 06 F 15/20	3 0 1	Q		
			G 06 F 15/20 3 0 1 H	
			G 06 F 15/20 3 0 1 Q	

審査請求 有

(全12頁)

(21)出願番号 特願昭61-72161
 (22)出願日 昭和61年(1986)3月29日

(71)出願人 000000307
 株式会社東芝
 神奈川県川崎市幸区堀川町72番地
 (72)発明者 岩井 勇
 神奈川県川崎市幸区小向東芝町1番地 株式会社東芝総合研究所内
 (72)発明者 土井 美和子
 神奈川県川崎市幸区小向東芝町1番地 株式会社東芝総合研究所内
 (72)発明者 岡本 利夫
 神奈川県川崎市幸区小向東芝町1番地 株式会社東芝総合研究所内
 (74)代理人 鈴江 武彦

(54)【発明の名称】文書レイアウト方法

(57)【要約】本公報は電子出願前の出願データであるため要約のデータは記録されません。

【特許請求の範囲】

(1) 入力された文書データを所定の区切りコードに従って1文毎に区切り、各文が見出し文であるか否かを判定する手段と、この判定結果に応じた文書構造規則に従って前記各文の文書構造をそれぞれ判定して前記文書データの階層的論理構造を求める手段と、この階層的論理構造の情報を前記文書データと対応付けて記憶する手段と、文書データの出力レイアウト形式を定めるレイアウト規則を前記各文書構造に対応して記憶したレイアウト規則辞書と、前記各文について求められた文書構造に応じたレイアウト規則に従ってその文の文書データを出力部に展開する手段とを具備したことを特徴とする文書処理装置。

(2) レイアウト規則辞書は、処理対象とする文書分野毎にその文書の階層構造に対する出力レイアウト形式を定める規則をそれぞれ格納したものである特許請求の範囲第1項記載の文書処理装置。

(3) レイアウト規則地所は、1つの文書構造に対応して複数のレイアウト規則を格納したものであり、これらの中から選択的にレイアウト規則を用いるものである特許請求の範囲第1項記載の文書処理装置。

(4) 文書の階層的論理構造は、章・節・箇条書き・段落等の文書論理構造からなるものである特許請求の範囲第1項記載の文書処理装置。

【発明の詳細な説明】

【発明の目的】

(産業上の利用分野)

本発明は文書データの文書構造を解析して求められる階層的論理構造に従って、その文書データの出力レイアウト形式を効果的に支援することのできる文書処理装置に関する。

(従来の技術)

ワードプロセッサ等の文書処理装置にあっては、文字コードや句読点コード等のコード情報の系列として文書データが人力される。そしてそのコード情報の系列で示される文書データを文書ファイルに登録したり、プリンタやディスプレイに出力したりする。

然し乍ら、コード情報で示される文字列をそのまま出力するだけでは、その文書が非常に読み難い。そこで一般的には、成るまどまりのある文の区切り位置に改行コードを挿入し、またその改行位置の次の文の先頭にスペース・コードを挿入して段落をつけたりして、その文書形式を整えることが行われる。

更には文書全体を、例えば章・節等の複数の範囲に分割し、そのまとまり毎に見出しを付したり、更にその文書を読み易くするべく、タブやインデント等の制御コードを挿入することも行われる。

ところが本来の文書データとは直接関係のない、例えば「二連した改行コード等の制御コードを挿入しながら文書作成することは文書作成の思考の妨げとなり、文書作

成効率の低下の原因となっている。

また、このようにして作成した文書を編集し直す場合には、上述した制御コードを削除したり、また別の箇所に挿入したりすることが必要となる。

この際、文書データに挿入される制御コードによって文書構造が変化することから、例えば数ページに亘る文書を再編集するような場合、その文書構造を全体的に統一するには多大な労力を必要とする等の問題があった。例えば文書形式を統一する為には、数ページ前の文書形式を参照する等の手続きが必要となる。これ故、簡易に文書の編集処理を進めることができず、その処理効率の向上を図ることが望めなかった。

(発明が解決しようとする問題点)

本発明はこのような事情を考慮してなされたもので、その目的とするところは、文書が持つ階層的論理構造を積極的に利用して文書処理を行なうことにより、上述した文書処理の煩わしさを解消して効果的な文書処理を可能ならしめる文書処理装置を提供することにある。

【発明の構成】

20 (問題点を解決するための手段)

本発明は、コード情報の系列として与えられる文書データに対して見出し語辞書および見出し規則辞書を参照して上記文書データ中から見出し文を抽出し、見出し文と見出いでない文とに分けられた各文について文書構造規則辞書を参照してその文書構造を判定し、この文書構造に従って前記文書データの、例えば章・節・箇条書き・段落等の文書構造で示される階層的論理構造を求め、この階層的論理構造によって示される各文の文書構造に対応したレイアウト規則に従って上記各文の文書データを30 それぞれ出力部に展開するようにしたものである。

つまり入力された文書データについて求められた階層的文書構造に従って、その文書を構成する各文の論理構造に対応するレイアウト規則に基き、その文書データをそれぞれ出力部に展開するようにしたものである。

(作用)

かくして本発明によれば、与えられた文書データの階層的論理構造に従って各文の文書データの出力部への展開構造、つまり出力レイアウト形式が自動的に決定され、その文書構造に応じたレイアウト形式でディスプレイ表示、またはプリント出力されることになる。

従って文書データの入力時、またはその編集処理時に一々その出力レイアウト形式を配慮することなく、つまり改行コードの挿入位置を配慮したり、インデントやタブの設定を行なうことなしにその文書処理することが可能となる。しかもこのようにして任意に文書処理しても、その文書データは予め定められたレイアウト形式に展開されてディスプレイ表示、またはプリント出力されることになる。

(実施例)

50 以下、図面を参照して本発明の実施例につき説明する。

第1図は実施例装置の概略構成図である。第1図において1は装置本体をなす文書管理部である。

キーボード等からなる入力部2を介してコード情報の系列として入力される文書データは、例えば第2図に示すような文書構造を持つものであり、上記文書管理部1の制御の下で原文記憶部3に格納され、文書処理に供される。そして文書管理部1にて文書処理された文書データは、表示制御部4の制御の下で表示部5にて表示されるようになっている。

第3図は入力文書データに対する文書構造解析の処理手続きを示すものであり、この流れに沿って本装置の機能を説明する。文書管理部1は前記入力部2から文書データを入力しくステップa)、これを上記原文記憶部3に格納すると共に、該文書データ中の区切りコード、例えば改行コードを検出し、この区切りコードにより区切られる1まとまりの文を順に抽出している。同時にその1文の長さを計測している。そして抽出した1文を単位として、以下に示す処理の実行を管理・制御している。

見出し抽出部6は、文書管理部1にて抽出された1文が見出しとしての可能性があるか否かを、上記の如く計測された1文の長さの情報と、見出し語辞書6aを参照して判定する(ステップb)。

この見出し語辞書6aは、見出しとして出現頻度の高い語句や記号を、例えば第4図に示すようにそのカテゴリ毎に分類して予め登録したものである。

具体的には見出しとして出現頻度の高い、例えば「はじめに」「あらすじ」等の語句を「見出し予約語」なるカテゴリにまとめて登録し、また見出しとして出現頻度の高い数字・記号を、それぞれのカテゴリ毎にまとめて登録している。

見出し抽出部6は、抽出された文の長さが所定の文字数(例えば40文字)以内であるか否かを判定し、所定の文字数以内である場合には見出しの可能性があると判定している。そしてこの文について、その文(コード情報の系列で示される語句や数字・記号)が見出し語辞書6aに登録されているか否かを検索しく(ステップb)、見出し語辞書Baに該当する語句が登録されている場合には、これを見出し候補としている(ステップC)。

しかして見出し判定部7は、その見出し規則辞書7aに格納されたm5図(a)～(d)に示す如き見出し規則に基いて、前記見出し抽出部6で見出し候補として抽出された文が上記見出し規則にマッチングするか否かを判定しく(ステップC)、その見出し候補が見出し規則にマッチングした場合、これを見出し文であると判定している(ステップe)。尚、前記見出し抽出部6にて見出し語が検出されなかった文、および見出し候補として判定された文であっても前記見出し規則に該当しなかった文は、見出し文でないと判定される(ステップf)。つまり段落等の文書本文であると判定する。

文書構造判定部8は、上記の如く見出し文、或いはそれ

以外の文として決定された各文に対して、文書構造規則辞書8aに格納された、例えば第6図(a)～(d)に示す如き文書構造規則に従い、その文が章見出しであるか、節見出しであるか、段落であるか等の文書論理構造をそれぞれ判定している(ステップg)。この文書構造判定部8にて判定された前記各文の論理構造を求める(ステップh)、その論理構造の情報を各文にそれぞれ対応付けて論理構造記憶部10に格納している(ステップi)。

10 尚、文書構造規則に基く論理構造解析に失敗した文については、その文の論理的構造に誤りがあるとして、例えば入力文書データの誤り修正を促す等のエラー処理が行われる(ステップj)。

文書展開部9は、以上の文書構造解析処理によって求められた入力文書データに対する階層的論理構造に従って、人力された各文をその論理構造に応じて定められたレイアウト形式で出力部に展開する。即ち、文書展開部9のレイアウト規則辞書9aには、その論理構造の種類に応じて定められた出力レイアウト形式を決定するレイアウト規則が格納されている。文書展開部9は、上述した如く各文についてそれぞれ求められた論理構造に従って、各文の文書データを前記表示部5にて表示する為のレイアウト規則を上記レイアウト規則辞書9aから求め、そのレイアウト規則に従って各文の文書データを前記表示制御部4にそれぞれ展開している。

前記表示制御部4は、このように展開された文書データの展開構造に従って前述した文書データの表示部5による表示を制御することになる。

かくしてこのように構成された装置によれば、次のようにして入力文書データの階層的論理構造が解析される。入力部2から文書データが入力されると、その文書データは原文記憶部3に順次格納されると共に、文書管理部1にて区切り処理される。この区切り処理は、人力コード情報が改行コードやスペース・コードか、或いは「・・・J r ; J r ; J等の区切り記号であるかを判定し、これらの区切りコードによって入力コード情報の系列を1文毎に切出すことによって行われる。この際、上記区切りコードによって区切られるコード情報の系列の長さを計数する等して、その文の長さ(文字数)が計測される。

しかして今、前記第2図に示す文書データが入力された場合について説明すると、改行コードによって区切られた第1行目の文「文書構造理解システム」、および第2行目の文「大川太郎」が与えられると、これらの各文について該当する語句が見出し語辞書6aに登録されていないことから、見出し抽出部6にてそれぞれ見出されないと判定される。文書構造判定部8はこの第1行目の文を、例えば文書の冒頭に出現する名詞句である等の規則に適合することから、その属性が標題であるとして判定する。また第2行目の文については、固有名詞、特に

人名を示す固有名詞であり、標題の後に出現する文である等の規則に従って著者名で、あると判定する。

しかる後、第3行目の文「1. はじめに」が与えられると、この文を構成している $r_1 J r$ 、 $J r$ はじめに 1 なる語句が前記見出し語辞書 6 a からそれぞれ見出される。この結果、この文は見出し候補 A として判定され、同時にその見出し候補を構成しているカテゴリが「(数字部) (後置部) (見出し予約語)」として求められる。すると見出し判定部 7 は、この見出し候補 A として判定された文の構造が見出し規則に適合しているか否かを見出し規則辞書 7 a を参照して調べる。

この見出しの判定処理は、先ず見出し候補 A を構成するカテゴリの並びを解析し、その解析構造が前記第5図に示す見出しとしての条件を満しているか否かを判定する。この場合、上記カテゴリ「(数字部) (後置部) (見出し予約語)」が、第5図 (a) ~ (d) に示す規則に従って第7図に示す如く解析され、見出しパターンを構成していることが確認されるから、見出し B であると判定される。尚、この判定処理によって上記カテゴリの並びが第5図に示す条件のいずれにも一致しないことが判定されたならば、上記見出し候補は見出しでないと決定されることになる。

しかして上記見出し B が求められると、次に前記文書構造判定部 8 にてその文書構造が第6図 (a) ~ (d) に例示する規則のどれに該当するかが判定される。この場合には、今までに分析された文の論理構造が前述したように「標題」と「著者名」であり、章見出しが出現していないことから第6図 (a) に示す見出し規則と照合される。この照合によって上記文「1. はじめに」なる見出し候補が第6図 (a) に示す規則の条件 (1) (1, 1) (1, 1, 1) (1, 1, 1, 1) に一致することが見出され、該見出し「1. はじめに」が章見出し C を構成していることが一意に決定される。そしてその論理構造の情報が、論、理構造記憶部 10 に格納される。しかる後、第4 ~ 5行目に亘る文が入力されると、その文字数が見出しとして可能性の成る所定の文字数を越えることから、見出し以外の文であると判定される。そしてこの場合には、その文が第6図 (c) に示す文書構造規則に該当していることから、段落を構成する文であると判定される。

以下、同様にして区切りコードによって区切られた文が見出し文であるか否かが判定され、文書構造規則辞書 8 a と照合されてその文書構造が順次求められる。

例えば見出しを構成する文が再び入力されると、先の例と同様にして見出し候補 A として検出され、第5図 (a) ~ (d) に示す見出し規則に適合することが判定される。そしてこの見出し B については、前記論理構造記憶部 10 の内容から既に章見出しが検出されていることが示されるので、先ず第6図 (b) に示す文書構造規則を参照し、上記見出し B のパターンがどの条件に該当する

かを調べる。そしてその見出し候補のパターンが条件 (1, 1) (2, 1) (3, 1) (4, 1) に該当する場合には、先に求められた章見出し C と同じレベルの見出しである可能性があることが判定される。しかる後、前記第6図 (a) に示す規則に適合するか否かを調べ、第6図 (a) に示す文書構造規則の条件 (1) (1, 1) (1, 1, 2) (1, 1, 2, 2) (1, 1, 2, 2, 1) に一致する場合には、その見出しを既に判定された先の章見出しと同一レベルの章見出しであると判定する。

尚、丸数字を冒頭に付した見出し文が与えられた場合には、それ以前に同様なパターンの見出しが検出されていないので第6図 (b) に示す規則 (条件) との照合にてマッチング不成功となる。

この結果、先に求められた見出しとはレベルの異なる見出しあることが判定される。その後、第6図 (a) に示す文書構造規則に対する照合により、例えば条件 (1) (1, 1) (1, 1, 2) (1, 1, 2, 2) (1, 1, 2, 2, 2) (1, 1, 2, 2, 2, 1) との一致が検出されてその見出し文が箇条書き見出しだると判定される。

尚、文が段落であるとして判定された場合には、その段落がどのレベルの見出しが受けたものかがわからない場合がある。このようなときには、例えば第6図 (d) に示す規則を参照して、段落と見出との接続関係を判定し、そのレベルを定めるようにすれば良い。

このような文書構造解析処理によって、その入力文書データを構成する各文の文書構造がそれぞれ求められ、その階層的論理構造の情報が入力文書データにそれぞれ対応して、例えば第8図に示すように論理構造記憶部 10 に格納される。

この第8図に示す文書の論理構造データについて簡単に説明すると、その文書構造の情報は、[]で囲んで示した文書データにそれぞれ対応付けて、()で囲んで示され5. 冒頭の数値にてその文の階層レベル、次の情報でその車馬性を、そして2で始まるデータにて文書データの解析結果を示している。この文書データの解析結果は、例えば次表のような意味を持っている。

このような意味を持つ記号の後に、その文書データが何であるか、或いは何文字目であるかの情報を付して、その解析結果が表現される。

さて前記文書展開部 9 は、上述した如く求められた文書データに対する階層的論理構造の情報に従い、該文書データを次のようにして出力部に展開している。

第9図は文書展開部 9 の 71-ドウエア構成を示すものである。論理構造データ読出し部 11 は、前記論理構造記憶部 10 に格納された前記第8図に示す如き文書データとその論理構造データを読出すもので、その文書データは1文字切出し部 12 にて1文字毎に切出されて文書データ割当て部 13 に与えられる。この際、文民性取出

し部14は、前記第8図に示すr t t l e J r a u t h e r J等の文民性データを抽出してレイアウト情報検出部15に与えており、またレイアウト情報解析部16は、前述した2の記号を付されて記述された論理構造情報を読み出している。

ここでレイアウト規則辞書9aには、例えば第10図(a) (b)に示すように、文書データを出力するフレーム(文書枠)に関する情報と、文民性に応じてその文をどのようにレイアウト出力するかの情報からなるレイアウト規則が格納されている。具体的には上記フレーム情報は、第10図(a)に示すようにディスプレイやプリンタ等の出力装置上でのフレーム管理番号、その出力位置の情報i、およびそのフレームの大きさの情報からなり、これによって文書データが出力される表示画面、または印刷用紙におけるフレームの構成が、例えば第11図に示す如く定義されるものとなっている。これらのフレーム情報は、文書データを出力するべく設定された複数のフレーム毎にそれぞれ定義される。

レイアウト規則読み出し部17は、このようなフレーム規則をレイアウト規則辞' #9aから読み出してフレーム形式情報バッファ18に格納すると共に、その情報を演算部19に与え、そのフレームの最大カラム数、および最大ライン数を求めている。この最大カラム数・ライン数のデータは最大カラム・ライン値バッファ20にセットされて後述する文字データの展開出力制御に用いられる。

また前記レイアウト規則辞書9aに格納された文民性に関するレイアウト規則は、例えば第10図(b)に示すように定義される。このレイアウト規則によって、その文民性の種別毎にその文書データを、前述した如く定義されるフレーム上にどのようにレイアウトするかが規定される。

ここでは、例えば「!!r」にて対象文字列の最後を強制改行すること、r c e Jにてその対象文字列をセンタリングすることがそれぞれ示される。

またr d s : I Jにて対象文書データを出力した後に1行分の改行を行なうことが示され、r u s : I Jにて対象文書データを出力する前に1行分の改行を行なうことが示される。更に「i S : 1」にてフレームの左側から1文字分のマージン(空白)を設けることが示される。尚、段落の場合にはその先頭行の1文字目を空白とする。

このようなレイアウト規則が前記レイアウト情報検出部15の制御の下で、その文民性の種別に従って上記レイアウト規則辞書9aから読み出されて前記レイアウト情報解析部16に与えられる。

レイアウト情報解析部18は、このようにして求められるレイアウト規則に基いて文書データ割当て部13を制御し、上記の如く1文字づつ切出される文書データを指定されたフレームに順に展開している。

即ち、レイアウト情報解析部16は、前記フレームの情報に従って文書データを展開するフレームを判定し、同時にそのフレームの大きさを判定している。そしてその指定されたフレームに、前記文民性に応じてレイアウト規則に従って、その文書データを展開制御している。

この際、1文字切出し部12にて文書データが1文字毎に切出される都度、これに同期してカラム・ラインカウンタ21の計数値がカウントアップされる。このカラム・ラインカウンタ21による計数値は、比較部22にて

10 前記最大カラム・ライン値バッファ20にセットされた値と逐次比較されており、その比較結果は判定部23にて判定されている。この判定部23にて前記1文字切出し部12にて切出された文字のレイアウト位置が、そのとき指定されているフレーム内であるか否かが判定される。

そしてレイアウト制御されて展開される文書データの割当て位置がその指定されたフレームを越えたことが判定されると、フレーム切換え部24が起動されて、前記文書データ割当て部13による文書データの割当てフレームが次のフレームに更新制御されるようになっている。

このようにして文書展開部9は、そのレイアウト規則辞1f9aに設定されているレイアウト規則に従って、その文書データを出力すべきフレームを設定している。そしてその設定したフレームに、上記文属性に応じて設定されたレイアウト規則に従ってその文書データを展開している。

例えば第8図に示す如く求められた論理構造の文書データにあっては、第1文が標題であり、そのレイアウト規則が「ノf J r c e J r d s ; I J」で示される

30 ことから、先ずレイアウト情報解析部16では前記最大カラム・ライン値バッファ20から、その設定されたフレームの最大カラム・ライン数を求め、例えば(MAXライン-対象文の文字数)+2等として、そのセンタリングして表示出力すべき文字位置を求めている。このデータによって前記カラム・ラインカウンタ21の計数値がプリセット的に変更され、その変更位置からその対象文の文書データの展開、つまり文字位置の割当てが行われる。

その後、対象文の文字データの出力割当てが終了した時40 点で、[d s ; L Jなる情報に従って強制改行され、出力された標題を示す文字列の下に1行-文の空白行が設定される。

このような文書データの出力(展開)制御が、前述した如く切出された各文毎に、その文について求められた論理構造に従って順次行われ、その文書データが所定のフレームに順に展開されていく。この結果、例えば第12図に示すように、前記第8図に示す如(階層的論理構造が求められた文書データが、それに対して指定されたレイアウト規則に従って出力展開されることになる。

50 かくしてこのように構成された装置によれば、人力文書

データからその文書の階層的論理（1 ■造が求められ、その階層的論理構造に従う文書データのレイアウト規則に従ってその文書データが、設定されたフレーム毎にその文属性に応じてそれぞれ定められたレイアウト形式で展開されることになる。

具体的には定められたフレームに文書データが順に割付けられ、またその文章の階層的論理構造に従う各文の文書構造の情報、つまりその文属性に従って、例えば強制改行やセンタリング等の制御が行われて文書データの展開が行われる。

従って文書作成時や文書編集時等の文書処理を、そのレイアウト形式を配慮することなしに文書内容のみに着目して効率良く実行することができる。

しかも、このようにして文書処理しても、上述した如く解析された階層的文書構造、およびその文書構造（文属性）によって指定されるレイアウト規則によって、その文書データの出力レイアウト形式が自動的に制御されるので、全体的に統一のとれた文書形式でその文書を作成することが可能となる。

従って文書の出力形式の制御の為に文書作成等の思考が妨げられることなく、その文書処理を簡単に、且つ効果的に行なうことが可能となる等の実用上多大なる効果が奏せられる。

尚、本発明は上述した実施例に限定されるものではない。例えば文書の論理構造を示すデータ形式や、その論理構造を求める為の規則、更には文書データの出力形式を定めるレイアウト規則は、その仕様に応じて定めれば良いものである。また字体（文字フォント）の変形を考慮した見出しの抽出と、文書構造の判定処理、更にはその出力制御を行なうようにしても良い。この場合、出力文字の高輝度表示や反転表示等を同時に制御することも勿論可能であり、そのような制御情報をレイアウト規則として設定しておけば十分である。

また上述した例では、文書データの階層的論理構造を求めた後、その出力レイアウト形式の制御を行なったが、1文毎にそのレイアウト展開処理を行なっても良い。この場合、その論理構造を一旦蓄積することなく、直接的にレイアウト展開処理に用いるようにしても良い。更には、1つの文民性について複数のレイアウト規則を準備しておき、その中から選択的にレイアウト規則を適用することも可能である。これによって同一文ぞきせいで異なる用途のレイアウト形式、例えば学会論文形式と社内報告書形式等に適宜変換することができ、種々のレイアウト編集を容易に行なうことが可能となる。またここでは日本語を例に説明したが、他国語に対しても同様に適用可能なことは勿論のことである。その他、本発明はその要旨を逸脱しない範囲で種々変形して実施することができる。

[発明の効果]

以」二説明したように本発明によれば、入力文書データ 50

からその文書の章・節・箇条書き・段落等の論理構造を求める、その論理構造に対応したレイアウト規則に従って文書データを所定の出力レイアウト形式で展開出力することができるので、文書出力形式の制御の為にその思考過程が妨げられることなく、その文書処理を効果的に進めることを可能ならしめる等の実用上多大なる効果が奏せられる。

【図面の簡単な説明】

図は本発明の一実施例を示すもので、第1図は実施例装置の概略構成図、第2図は入力文書データの一例を示す図、第3図は実施例装置における文書構造解析処理の手続きの流れを示す図、第4図は見出し語辞書の一例を示す図、第5図(a)～(d)は見出し抽出規則辞書の構成例を示す図、第6図(a)～(d)は文書構造規則辞書の構成例を示す図、第7図は見出し文の解析構造を示す図、第8図は論理構造記憶部に格納された文書構造の情報の例を示す図、第9図は文書展開部の構成例を示す図、第10図はレイアウト規則辞書の構成例を示す図、第11図はレイアウト規則によって示されるフレーム(文書枠)の例を示す図、第12図はレイアウト展開処理されて出力される文書形式の例を示す図である。

■…文書管理部、2…入力部、3…原文記憶部、4…表示制御部、5…表示部、6…見出し抽出部、6a…見出し語辞書、7…見出し判定部、7a…見出し規則辞書、8…文書構造判定部、8a…文書構造規則部、9…文書展開部、9a…レイアウト規則辞書、10…論理構造記憶部。

出願人代理人 弁理士 鈴江武彦

30 第2図

(a)

第6図

(d)

第6図

第7図

■ 8 図

第10図

第11図

第12図

⑩ 日本国特許庁 (JP) ⑪ 特許出願公開
 ⑫ 公開特許公報 (A) 昭62-229364

⑬ Int.Cl.⁴
 G 06 F 15/20

識別記号
 301

序内整理番号
 H-7218-5B
 Q-7218-5B

⑭ 公開 昭和62年(1987)10月8日

審査請求 未請求 発明の数 1 (全12頁)

⑮ 発明の名称 文書処理装置

⑯ 特 願 昭61-72161
 ⑰ 出 願 昭61(1986)3月29日

⑱ 発明者 岩井 勇 川崎市幸区小向東芝町1番地 株式会社東芝総合研究所内
 ⑲ 発明者 土井 美和子 川崎市幸区小向東芝町1番地 株式会社東芝総合研究所内
 ⑳ 発明者 岡本 利夫 川崎市幸区小向東芝町1番地 株式会社東芝総合研究所内
 ㉑ 出願人 株式会社 東芝 川崎市幸区堀川町72番地
 ㉒ 代理人 井理士 鈴江 武彦 外2名

明細書

1. 発明の名称

文書処理装置

2. 特許請求の範囲

(1) 入力された文書データを所定の区切りコードに従って1文毎に区切り、各文が見出し文であるか否かを判定する手段と、この判定結果に応じた文書構造規則に従って前記各文の文書構造をそれぞれ判定して前記文書データの階層的論理構造を求める手段と、この階層的論理構造の情報を前記文書データと対応付けて記憶する手段と、文書データの出力レイアウト形式を定めるレイアウト規則を前記各文書構造に対応して記憶したレイアウト規則辞書と、前記各文について求められた文書構造に応じたレイアウト規則に従ってその文の文書データを出力部に展開する手段とを具備したことを特徴とする文書処理装置。

(2) レイアウト規則辞書は、処理対象とする文書分野毎にその文書の階層構造に対する出力レイアウト形式を定める規則をそれぞれ格納したもので

ある特許請求の範囲第1項記載の文書処理装置。

(3) レイアウト規則辞書は、1つの文書構造に対応して複数のレイアウト規則を格納したものであり、これらの中から選択的にレイアウト規則を用いるものである特許請求の範囲第1項記載の文書処理装置。

(4) 文書の階層的論理構造は、章・節・箇条書き・段落等の文書論理構造からなるものである特許請求の範囲第1項記載の文書処理装置。

3. 発明の詳細な説明

【発明の目的】

(産業上の利用分野)

本発明は文書データの文書構造を解析して求められる階層的論理構造に従って、その文書データの出力レイアウト形式を効率的に支援することができる文書処理装置に関する。

(従来の技術)

ワードプロセッサ等の文書処理装置にあっては、文字コードや句終点コード等のコード情報の系列として文書データが入力される。そしてその

特開昭62-229364 (2)

コード情報の系列で示される文書データを文書ファイルに登録したり、プリンタやディスプレイに出力したりする。

然し乍ら、コード情報で示される文字列をそのまま出力するだけでは、その文書が非常に読み難い。そこで一般的には、或るまとまりのある文の区切り位置に改行コードを挿入し、またその改行位置の次の文の先頭にスペース・コードを挿入して段落をつけたりして、その文書形式を整えることが行われる。

更には文書全体を、例えば章・節等の複数の範囲に分割し、そのまとまり毎に見出しを付したり、更にその文書を読み易くするべく、タブやインデント等の制御コードを挿入することも行われる。

ところが本来の文書データとは直接関係のない、例えば上述した改行コード等の制御コードを挿入しながら文書作成することは文書作成の思考の妨げとなり、文書作成効率の低下の原因となっている。

また、このようにして作成した文書を編集し直

す場合には、上述した制御コードを削除したり、また別の箇所に挿入したりすることが必要となる。この際、文書データに挿入される制御コードによって文書構造が変化することから、例えば数ページに亘る文書を再編集するような場合、その文書構造を全体的に統一するには多大な労力を必要とする等の問題があった。例えば文書形式を統一するには、数ページ前の文書形式を参照する等の手続きが必要となる。これ故、簡易に文書の編集処理を進めることができず、その処理効率の向上を図ることが望めなかった。

(発明が解決しようとする問題点)

本発明はこのような事情を考慮してなされたもので、その目的とするとところは、文書が持つ階層的論理構造を積極的に利用して文書処理を行なうことにより、上述した文書処理の煩わしさを解消して効率的な文書処理を可能ならしめる文書処理装置を提供することにある。

【発明の構成】

(問題点を解決するための手段)

本発明は、コード情報の系列として与えられる文書データに対して見出し語辞書および見出し規則辞書を参照して上記文書データ中から見出し文を抽出し、見出し文と見出いでない文とに分けられた各文について文書構造規則辞書を参照してその文書構造を判定し、この文書構造に従って前記文書データの、例えば章・節・箇条書き・段落等の文書構造で示される階層的論理構造を求め、この階層的論理構造によって示される各文の文書構造に対応したレイアウト規則に従って上記各文の文書データをそれぞれ出力部に展開するようにしたものである。

つまり入力された文書データについて求められた階層的文書構造に従って、その文書を構成する各文の論理構造に対応するレイアウト規則に基き、その文書データをそれぞれ出力部に展開するようにしたものである。

(作用)

かくして本発明によれば、与えられた文書データの階層的論理構造に従って各文の文書データの出力部への展開構造、つまり出力レイアウト形式が自動的に決定され、その文書構造に応じたレイアウト形式でディスプレイ表示、またはプリント出力されることになる。

従って文書データの入力時、またはその編集処理時に一々その出力レイアウト形式を記述することなく、つまり改行コードの挿入位置を記述したり、インデントやタブの設定を行なうことなしにその文書処理することが可能となる。しかもこのようにして任意に文書処理しても、その文書データは予め定められたレイアウト形式に展開されてディスプレイ表示、またはプリント出力されることになる。

(実施例)

以下、図面を参照して本発明の実施例について説明する。

第1図は実施例装置の概略構成図である。第1

特開昭62-229364 (3)

図において 1は装置本体をなす文書管理部である。キーボード等からなる入力部2を介してコード情報の系列として入力される文書データは、例えば第2図に示すような文書構造を持つものであり、上記文書管理部1の制御の下で原文記憶部3に格納され、文書処理に供される。そして文書管理部1にて文書処理された文書データは、表示制御部4の制御の下で表示部5にて表示されるようになっている。

第3図は入力文書データに対する文書構造解析の処理手続を示すものであり、この流れに沿って本装置の機能を説明する。文書管理部1は前記入力部2から文書データを入力し(ステップa)、これを上記原文記憶部3に格納すると共に、該文書データ中の区切りコード、例えば改行コードを検出し、この区切りコードにより区切られる1まとまりの文を順に抽出している。同時にその1文の長さを計測している。そして抽出した1文を単位として、以下に示す処理の実行を管理・制御している。

れを見出し候補としている(ステップc)。

しかして見出し判定部7は、その見出し規則辞書7aに格納された第5図(a)～(d)に示す如き見出し規則に基いて、前記見出し抽出部8で見出し候補として抽出された文が上記見出し規則にマッチングするか否かを判定し(ステップc)、その見出し候補が見出し規則にマッチングした場合、これを見出し文であると判定している(ステップe)。尚、前記見出し抽出部8にて見出し候補が検出されなかった文、および見出し候補として判定された文であっても前記見出し規則に該当しなかった文は、見出し文でないと判定される(ステップf)。つまり段落等の文書本文であると判定する。

文書構造判定部8は、上記の如く見出し文、或いはそれ以外の文として決定された各文に対して、文書構造規則辞書8aに格納された、例えば第6図(a)～(d)に示す如き文書構造規則に従い、その文が章見出しであるか、節見出しであるか、段落であるか等の文書論理構造をそれぞれ判定し

見出し抽出部8は、文書管理部1にて抽出された1文が見出しとしての可能性があるか否かを、上記の如く計測された1文の長さの情報と、見出し語辞書6aを参照して判定する(ステップg)。この見出し語辞書6aは、見出しとして出現頻度の高い語句や記号を、例えば第4図に示すようにそのカテゴリ毎に分類して予め登録したものである。具体的には見出しとして出現頻度の高い、例えば「はじめに」「あらすじ」等の語句を「見出し予約語」なるカテゴリにまとめて登録し、また見出しとして出現頻度の高い数字・記号を、それぞれのカテゴリ毎にまとめて登録している。

見出し抽出部8は、抽出された文の長さが所定の文字数(例えば40文字)以内であるか否かを判定し、所定の文字数以内である場合には見出しの可能性があると判定している。そしてこの文について、その文(コード情報の系列で示される語句や数字・記号)が見出し語辞書6aに登録されているか否かを検索し(ステップh)、見出し語辞書6aに該当する語句が登録されている場合には、こ

ている(ステップg)。この文書構造判定部8にて判定された前記各文の論理構造を求める(ステップh)、その論理構造の情報を各文にそれぞれ対応付けて論理構造記憶部10に格納している(ステップi)。

尚、文書構造規則に基く論理構造解析に失敗した文については、その文の論理的構造に誤りがあるとして、例えば入力文書データの誤り修正を促す等のエラー処理が行われる(ステップj)。

文書展開部9は、以上の文書構造解析処理によって求められた入力文書データに対する階層的論理構造に従って、入力された各文をその論理構造に応じて定められたレイアウト形式で出力部に展開する。即ち、文書展開部9のレイアウト展開辞書9aには、その論理構造の種類に応じて定められた出力レイアウト形式を決定するレイアウト規則が格納されている。文書展開部9は、上述した如く各文についてそれぞれ求められた論理構造に従って、各文の文書データを前記表示部5にて表示する為のレイアウト規則を上記レイアウト規則辞書9aに記載された出力レイアウト形式に従って、各文の文書データを前記表示部5にて表示する。

特開昭62-229364 (4)

書5aから求め、そのレイアウト規則に従って各文の文書データを前記表示制御部4にそれぞれ展開している。

前記表示制御部4は、このように展開された文書データの展開構造に従って前述した文書データの表示部5による表示を制御することになる。

かくしてこのように構成された装置によれば、次のようにして入力文書データの階層的論理構造が解析される。

入力部2から文書データが入力されると、その文書データは原文記憶部3に順次格納されると共に、文書管理部1にて区切り処理される。この区切り処理は、入力コード情報が改行コードやスペース・コードか、或いは「…」「;」「:」等の区切り記号であるかを判定し、これらの区切りコードによって入力コード情報の系列を1文毎に切出すことによって行われる。この際、上記区切りコードによって区切られるコード情報の系列の長さを計数する等して、その文の長さ（文字数）が計測される。

（見出し予約語）として求められる。

すると見出し判定部7は、この見出し候補Aとして判定された文の構造が見出し規則に適合しているか否かを見出し規則辞書7aを参照して調べる。この見出しの判定処理は、先ず見出し候補Aを構成するカテゴリの並びを解析し、その解析構造が前記第5図に示す見出しとしての条件を満しているか否かを判定する。この場合、上記カテゴリ「（数字部）（後置部）（見出し予約語）」が、第5図（a）～（d）に示す規則に従って第7図に示す如く解析され、見出しパターンを構成していることが確認されるから、見出しBであると判定される。尚、この判定処理によって上記カテゴリの並びが第5図に示す条件のいずれにも一致しないことが判定されたならば、上記見出し候補は見出でないと決定されることになる。

しかして上記見出しBが求められると、次に前記文書構造判定部8にてその文書構造が第6図（a）～（d）に示す規則のどれに該当するかが判定される。この場合には、今までに分析さ

しかして今、前記第2図に示す文書データが入力された場合について説明すると、改行コードによって区切られた第1行目の文「文書構造理解システム」、および第2行目の文「大川太郎」が与えられると、これらの各文については該当する語句が見出し語辞書6aに登録されていないことから、見出し抽出部6にてそれぞれ見出でないと判定される。文書構造判定部8はこの第1行目の文を、例えば文書の冒頭に出現する名詞句である等の規則に適合することから、その属性が標題であるとして判定する。また第2行目の文については、固有名詞、特に人名を示す固有名詞であり、標題の後に出現する文である等の規則に従って著者名であると判定する。

しかる後、第3行目の文「1.はじめに」が与えられると、この文を構成している「1」「.」「はじめに」なる語句が前記見出し語辞書6aからそれぞれ見出される。この結果、この文は見出し候補Aとして判定され、同時にその見出し候補を構成しているカテゴリが「（数字部）（後置部）

れた文の論理構造が前述したように「標題」と「著者名」であり、章見出しが出現していないことから第6図（a）に示す見出し規則と照合される。この照合によって上記文「1.はじめに」なる見出し候補が第6図（a）に示す規則の条件（1）（1, 1）（1, 1, 1）（1, 1, 1, 1）に一致することが見出され、該見出し「1.はじめに」が章見出しCを構成していることが一意に決定される。そしてその論理構造の情報が、論理構造記憶部10に格納される。

しかる後、第4～5行目に亘る文が入力されると、その文字数が見出しとして可能性の或る所定の文字数を越えることから、見出し以外の文であると判定される。そしてこの場合には、その文が第6図（c）に示す文書構造規則に該当していることから、段落を構成する文であると判定される。

以下、同様にして区切りコードによって区切られた文が見出し文であるか否かが判定され、文書構造規則辞書8aと照合されてその文書構造が順次求められる。

特開昭62-229364 (5)

例えば見出しを構成する文が再び入力されると、先の例と同様にして見出し候補Aとして検出され、第5図(a)～(d)に示す見出し規則に適合することが判定される。そしてこの見出しBについては、前記論理構造記憶部10の内容から既に章見出しが検出されていることが示されるので、先ず第6図(b)に示す文書構造規則を参照し、上記見出しBのパターンがどの条件に該当するかを調べる。そしてその見出し候補のパターンが条件(1, 1) (2, 1) (3, 1) (4, 1)に該当する場合には、先に求められた章見出しCと同じレベルの見出しである可能性があることが判定される。しかる後、前記第6図(a)に示す規則に適合するか否かを調べ、第6図(a)に示す文書構造規則の条件(1) (1, 1) (1, 1, 2) (1, 1, 2, 2) (1, 1, 2, 2, 1)に一致する場合には、その見出しを既に判定された先の章見出しと同一レベルの章見出しであると判定する。

尚、丸数字を冒頭に付した見出し文が与えられ

書データにそれぞれ対応して、例えば第8図に示すように論理構造記憶部10に格納される。

この第8図に示す文書の論理構造データについて簡単に説明すると、その文書構造の情報は、〔〕で囲んで示した文書データにそれぞれ対応付けて、〔〕で囲んで示され、冒頭の数値にてその文の階層レベル、次の情報でその文属性を、そしてzで始まるデータにて文書データの解析結果を示している。この文書データの解析結果は、例えば次表のような意味を持っている。

た場合には、それ以前に同様なパターンの見出しが検出されていないので第6図(b)に示す規則(条件)との組合にてマッチング不成功となる。この結果、先に求められた見出しとはレベルの異なる見出しであることが判定される。その後、第6図(a)に示す文書構造規則に対する組合により、例えば条件(1) (1, 1) (1, 1, 2) (1, 1, 2, 2) (1, 1, 2, 2, 2) (1, 1, 2, 2, 1)との一致が検出されてその見出し文が箇条書き見出しであると判定される。

尚、文が段落であるとして判定された場合には、その段落がどのレベルの見出しを受けたものかがわからない場合がある。このようなときには、例えば第6図(d)に示す規則を参照して、段落と見出しとの接続関係を判定し、そのレベルを定めるようにすれば良い。

このような文書構造解析処理によって、その入力文書データを構成する各文の文書構造がそれぞれ求められ、その階層的論理構造の情報が入力文

記号	意味
z 1	記号部
z 2	英数字括弧始め
z 3	予約語部
z 4	英数字部1
z 5	英数字部2
z 6	英数字括弧終り
z 7	後尾部1
{ }	{ }
z 21	見出し予約語
z 22	主見出し本体始め位置
z 23	主見出し本体終り位置
z 24	後見出し記号
{ }	{ }
z 28	内容部始め位置
z 29	内容部終り位置
{ }	{ }

このような意味を持つ記号の後に、その文書データが何であるか、或いは何文字目であるかの情報を付して、その解析結果が表現される。

特開昭62-229364 (6)

さて前記文書展開部9は、上述した如く求められた文書データに対する階層的論理構造の情報を従い、該文書データを次のようにして出力部に展開している。

第9図は文書展開部9のハードウェア構成を示すものである。論理構造データ読出し部11は、前記論理構造記憶部10に格納された前記第8図に示す如き文書データとその論理構造データを読み出すもので、その文書データは1文字切出し部12にて1文字毎に切出されて文書データ割当て部13に与えられる。この際、文属性取出し部14は、前記第8図に示す「title」「author」等の文属性データを抽出してレイアウト情報検出部15に与えており、またレイアウト情報解析部16は、前述したzの記号を付されて記述された論理構造情報を読み出している。

ここでレイアウト規則辞書9aには、例えば第10図(a)(b)に示すように、文書データを出力するフレーム(文書枠)に関する情報と、文属性に応じてその文をどのようにレイアウト出力

属性に関するレイアウト規則は、例えば第10図(b)に示すように定義される。このレイアウト規則によって、その文属性の種別毎にその文書データを、前述した如く定義されるフレーム上にどのようにレイアウトするかが規定される。

ここでは、例えば「*ss*」にて対象文字列の最後を強制改行すること、「*Co*」にてその対象文字列をセンタリングすることがそれぞれ示される。また「*ds;1*」にて対象文書データを出力した後に1行分の改行を行なうことが示され、「*us;1*」にて対象文書データを出力する前に1行分の改行を行なうことが示される。更に「*ss;1*」にてフレームの左側から1文字分のマージン(空白)を設けることが示される。尚、段落の場合にはその先頭行の1文字目を空白とする。

このようなレイアウト規則が前記レイアウト情報検出部15の制御の下で、その文属性の種別に従って上記レイアウト規則辞書9aから読み出されて前記レイアウト情報解析部16に与えられる。

レイアウト情報解析部16は、このようにして求

するかの情報からなるレイアウト規則が格納されている。具体的には上記フレーム情報は、第10図(e)に示すようにディスプレイヤやプリンタ等の出力装置上のフレーム管理番号、その出力位置の情報、およびそのフレームの大きさの情報からなり、これによって文書データが出力される表示画面、または印刷用紙におけるフレームの構成が、例えば第11図に示す如く定義されるものとなっている。これらのフレーム情報は、文書データを出力するべく設定された複数のフレーム毎にそれぞれ定義される。

レイアウト規則読出し部17は、このようなフレーム規則をレイアウト規則辞書9aから読み出してフレーム形式情報バッファ18に格納すると共に、その情報を演算部19に与え、そのフレームの最大カラム数、および最大ライン数を求めている。この最大カラム数・ライン数のデータは最大カラム・ライン値バッファ20にセットされて後述する文字データの展開出力制御に用いられる。

また前記レイアウト規則辞書9aに格納された文

められるレイアウト規則に基いて文書データ割当て部13を制御し、上記の如く1文字づつ切出される文書データを指定されたフレームに順に展開している。

即ち、レイアウト情報解析部16は、前記フレームの情報に従って文書データを展開するフレームを判定し、同時にそのフレームの大きさを判定している。そしてその指定されたフレームに、前記文属性に応じてレイアウト規則に従って、その文書データを展開制御している。

この際、1文字切出し部12にて文書データが1文字毎に切出される毎度、これに同期してカラム・ラインカウンタ21の計数値がカウントアップされる。このカラム・ラインカウンタ21による計数値は、比較部22にて前記最大カラム・ライン値バッファ20にセットされた値と逐次比較されており、その比較結果は判定部23にて判定されている。この判定部23にて前記1文字切出し部12にて切出された文字のレイアウト位置が、そのとき指定されているフレーム内であるか否かが判定される。

特開昭62-229364 (7)

そしてレイアウト制御されて展開される文書データの割当て位置がその指定されたフレームを越えたことが判定されると、フレーム切換部24が起動されて、前記文書データ割当て部13による文書データの割当てフレームが次のフレームに更新制御されるようになっている。

このようにして文書展開部9は、そのレイアウト規則辞書9aに設定されているレイアウト規則に従って、その文書データを出力すべきフレームを設定している。そしてその設定したフレームに、上記文属性に応じて設定されたレイアウト規則に従ってその文書データを展開している。

例えば第8図に示す如く求められた論理構造の文書データにあっては、第1文が標題であり、そのレイアウト規則が「主」「Ce」「ds;1」で示されることから、先ずレイアウト情報解析部10では前記最大カラム・ライン数バッファ20から、その設定されたフレームの最大カラム・ライン数を求める。例えば

(MAXライン-対象文の文字数) + 2

求められ、その階層的論理構造に従う文書データのレイアウト規則に従ってその文書データが、設定されたフレーム毎にその文属性に応じてそれぞれ定められたレイアウト形式で展開されることになる。

具体的には定められたフレームに文書データが順に割付けられ、またその文章の階層的論理構造に従う各文の文書構造の情報、つまりその文属性に従って、例えば強制改行やセンタリング等の制御が行われて文書データの展開が行われる。

従って文書作成時や文書編集時等の文書処理を、そのレイアウト形式を配慮することなしに文書内容のみに着目して効率良く実行することができる。しかも、このようにして文書処理しても、上述した如く解析された階層的文書構造、およびその文書構造(文属性)によって指定されるレイアウト規則によって、その文書データの出力レイアウト形式が自動的に制御されるので、全体的に統一のとれた文書形式でその文書を作成することが可能となる。

等として、そのセンタリングして表示出力すべき文字位置を求めている。このデータによって前記カラム・ラインカウンタ21の計数値がプリセット的に変更され、その変更位置からその対象文の文書データの展開、つまり文字位置の割当てが行われる。

その後、対象文の文字データの出力割当てが終了した時点で、「ds;1」なる情報に従って強制改行され、出力された標題を示す文字列の下に1行文の空白行が設定される。

このような文書データの出力(展開)制御が、前述した如く切出された各文毎に、その文について求められた論理構造に従って順次行われ、その文書データが所定のフレームに順に展開されていく。この結果、例えば第12図に示すように、前記第8図に示す如く階層的論理構造が求められた文書データが、それに対して指定されたレイアウト規則に従って出力展開されることになる。

かくしてこのように構成された装置によれば、入力文書データからその文書の階層的論理構造が

従って文書の出力形式の制御の為に文書作成等の思考が妨げられることなく、その文書処理を簡易に、且つ効率的に行なうことが可能となる等の実用上多大なる効果が発せられる。

尚、本発明は上述した実施例に限定されるものではない。例えば文書の論理構造を示すデータ形式や、その論理構造を求める為の規則、更には文書データの出力形式を定めるレイアウト規則は、その仕様に応じて定めれば良いものである。また字体(文字フォント)の変形を考慮した見出しの抽出と、文書構造の判定処理、更にはその出力制御を行なうようにしても良い。この場合、出力文字の高輝度表示や反転表示等を同時に制御することも勿論可能であり、そのような制御情報をレイアウト規則として設定しておけば十分である。

また上述した例では、文書データの階層的論理構造を求めた後、その出力レイアウト形式の制御を行なったが、1文毎にそのレイアウト展開処理を行なっても良い。この場合、その論理構造を一旦蓄積することなく、直接的にレイアウト展開処

理に用いるようにしても良い。更には、1つの文属性について複数のレイアウト規則を準備しておき、その中から選択的にレイアウト規則を適用することも可能である。これによって同一文だけをいで異なる用途のレイアウト形式、例えば学会論文形式と社内報告書形式等に適宜変換することができ、種々のレイアウト編集を容易に行うことが可能となる。またここでは日本語を例に説明したが、他の国語に対しても同様に通用可能なことは勿論のことである。その他、本発明はその要旨を逸脱しない範囲で種々変形して実施することができる。

【発明の効果】

以上説明したように本発明によれば、入力文書データからその文書の章・節・箇条書き・段落等の論理構造を求め、その論理構造に対応したレイアウト規則に従って文書データを所定の出力レイアウト形式で展開出力することができるので、文書出力形式の制御の為にその思考過程が妨げられることなく、その文書処理を効果的に進められることがなく、その文書処理を効果的に進める

6a…見出し語辞書、7…見出し判定部、7a…見出し規則辞書、8…文書構造判定部、8a…文書構造規則部、9…文書展開部、9a…レイアウト規則辞書、10…論理構造記憶部。

出願人代理人 弁理士 神江武吉

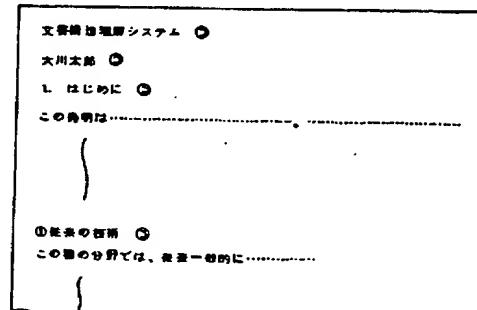
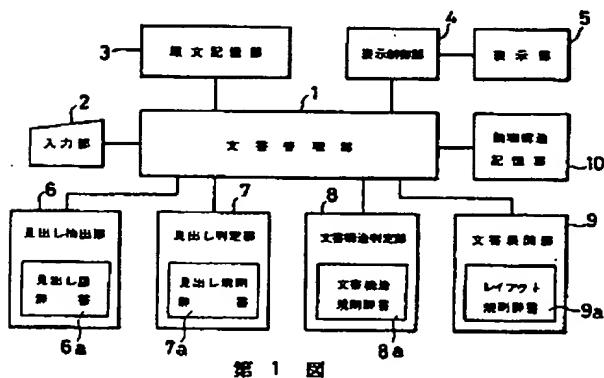
特開昭62-229364 (8)

ことを可能ならしめる等の実用上多大なる効果が奏せられる。

4. 図面の簡単な説明

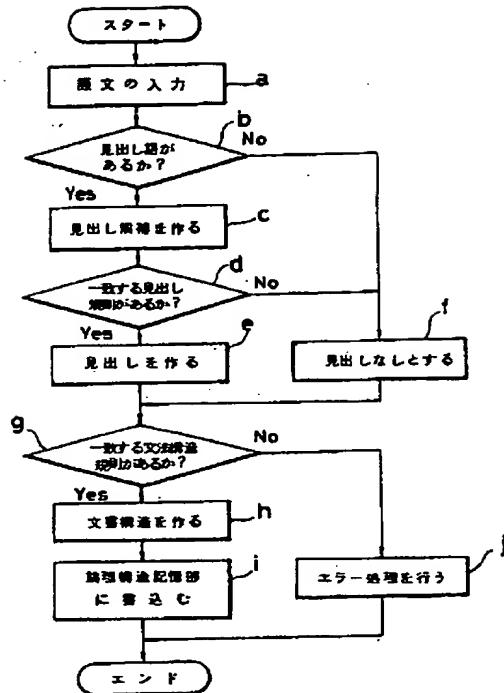
図は本発明の一実施例を示すもので、第1図は実施例装置の概略構成図、第2図は入力文書データの一例を示す図、第3図は実施例装置における文書構造解析処理の手続きの流れを示す図、第4図は見出し語辞書の一例を示す図、第5図(a)～(d)は見出し抽出規則辞書の構成例を示す図、第6図(a)～(d)は文書構造規則辞書の構成例を示す図、第7図は見出し文の解析構造を示す図、第8図は論理構造記憶部に格納された文書構造の情報の例を示す図、第9図は文書展開部の構成例を示す図、第10図はレイアウト規則辞書の構成例を示す図、第11図はレイアウト規則によって示されるフレーム(文書枠)の例を示す図、第12図はレイアウト展開処理されて出力される文書形式の例を示す図である。

1…文書管理部、2…入力部、3…原文記憶部、4…表示制御部、5…表示部、6…見出し抽出部、



第2図

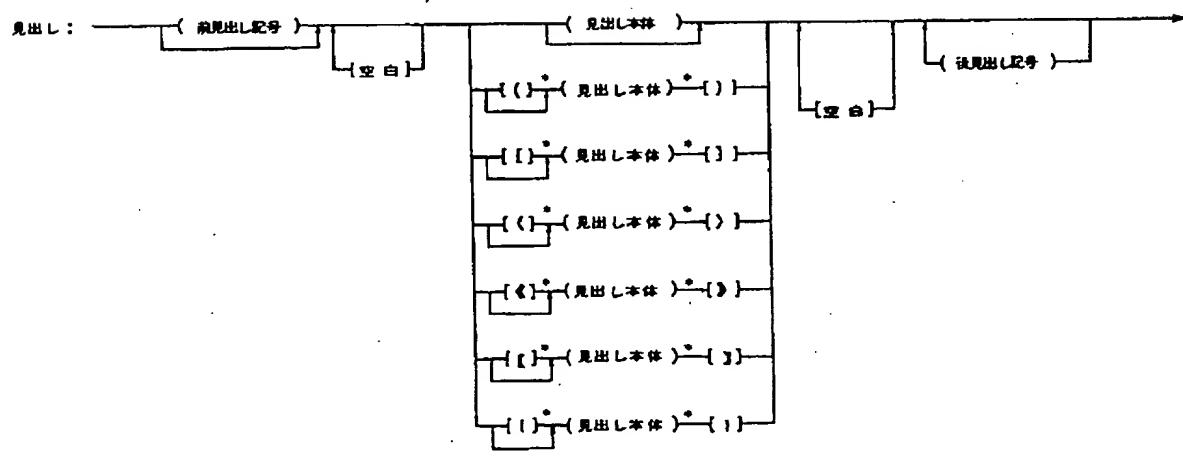
特開昭62-229364 (9)



第 3 图

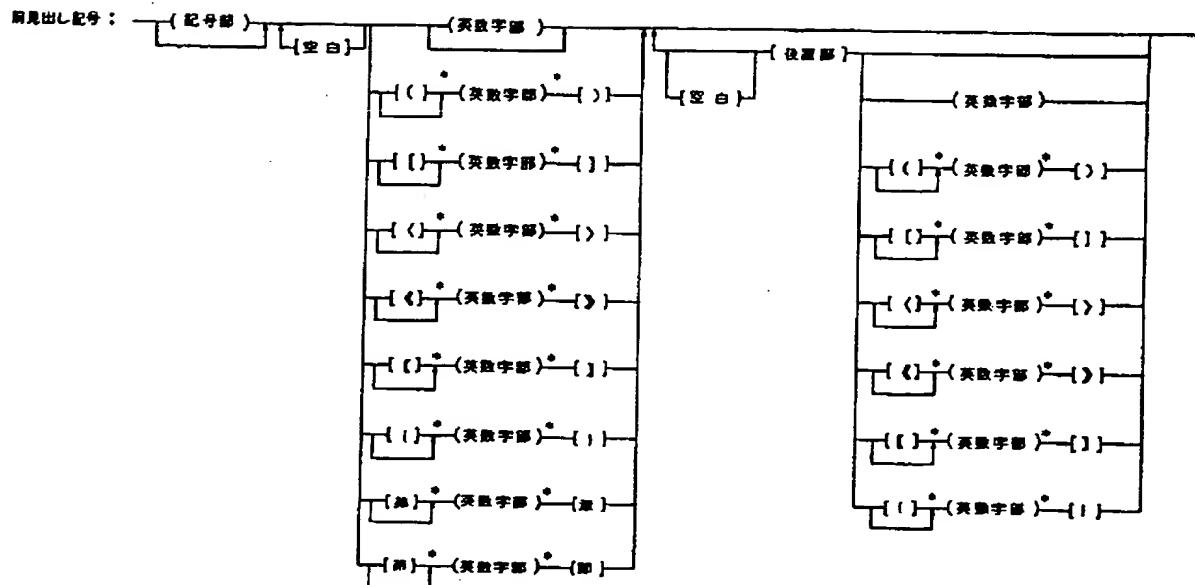
カタゴリ		検索項目
見出し予約語		はじめ あらすじ }
記号部		S 六 一
数字部	一般	1 2 3
	九数字	① ② ③
	ローマ数字	I II III
		1
後見出し記号		...
		2
後編部		• :
		3

第 4 圖

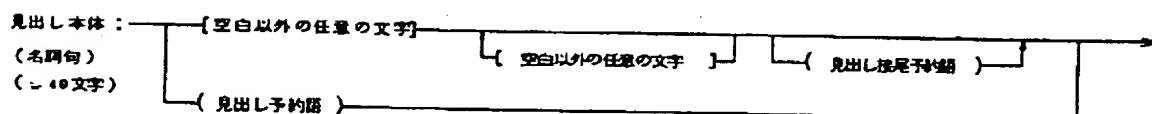


(a)

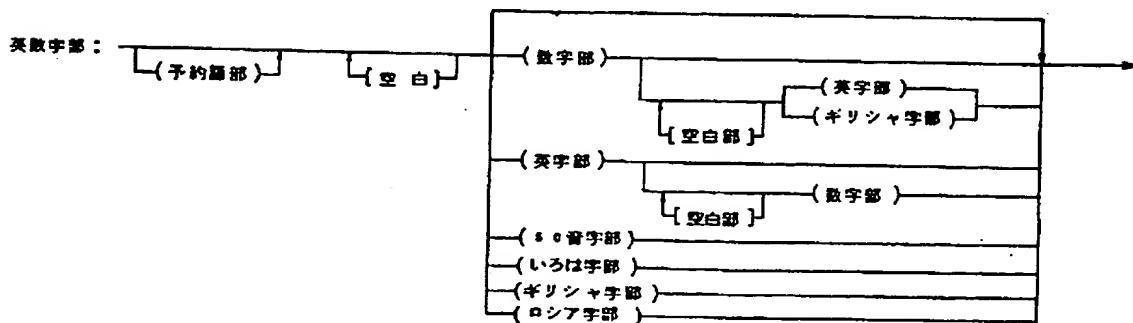
特圖昭 62-229364 (10)



(b) 第 5 國



(c)



(d)

特開昭62-229364 (11)

【見出しに該する場所】

条件1 見出し部を含まない
 条件1.1 見出し本体を含む
 条件1.1.1 見出し子部を含む
 条件1.1.1.1 斜方に見出しがない
 (細目) 見出しがある
 ~ 記号部, 英数字部, 文字部, 見出し記号を
 見出しそうとする

条件1.1.2 見出し子部を含まない
 条件1.1.2.1 斜方に見出しがある
 条件1.1.2.1.1 見出しそうとするとマッチングに成功
 (細目) 見出しがある
 ~ 斜見出しそうとするのオーダを1増す
 条件1.1.2.2.1.1 斜方に見出しがない
 条件1.1.2.2.1.1.1 見出しそうとするとマッチングに成功
 (細目) 見出しそうとする
 条件1.1.2.2.2.1.1 斜方に見出しそうとする
 条件1.1.2.2.2.2.1.1 見出しそうとするとマッチングに成功
 (細目) 今までの見出しそうとするオーダを1増す

(a)

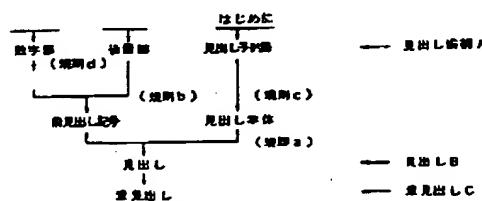
第6図

【見出しパターンとのマッチングに関する規則】

条件1.1 英数字部を含んでいる
 条件1.1 記号部, 英数字部の字體が同じである
 条件1.1.1 英数字部のオーダが見出しそうとするオーダの+1である
 条件1.1.1.1 斜号部, 英数字部のオーダ以外の部分, 文字部, 見出し記号, 見出しほど本体の字體の有無が見出しそうとすると同一である
 (細目) マッチング成功
 条件1.1.2 記号部, 英数字部のオーダ以外の部分, 文字部, 見出し記号, 見出しほど本体の字體の有無が見出しそうとすると同一でない
 (細目) マッチング失敗
 条件1.1.2.1 英数字部のオーダが見出しそうとするオーダの+1か+2である
 条件1.1.2.1.1 斜号部, 英数字部のオーダ以外の部分, 文字部, 見出し記号, 見出しほど本体の字體の有無が見出しそうとすると同一である
 (細目) マッチング成功
 条件1.1.2.2 記号部, 英数字部のオーダ以外の部分, 文字部, 見出し記号, 見出しほど本体の字體の有無が見出しそうとすると同一でない
 (細目) マッチング失敗
 条件1.1.2.3 記号部を含んでいない
 条件1.1.3 記号部, 英数字部のオーダ以外の部分, 文字部, 見出し記号, 見出しほど本体の字體の有無が見出しそうとする
 (細目) マッチング成功
 条件1.1.3.1 記号部, 英数字部のオーダ以外の部分, 文字部, 見出し記号, 見出しほど本体の字體の有無が見出しそうとすると同一でない
 (細目) マッチング失敗

(b)

第6図



第7図

【段落に関する規則】

条件1.1 見出しが含んでいない
 (細目) 整流である

(c)

【表示範囲に関する規則】

条件1.1 整流である
 条件1.1.1 目次が見出しだす
 (細目) 1つ目や見出しそうとするオーダーと同一レベルにする

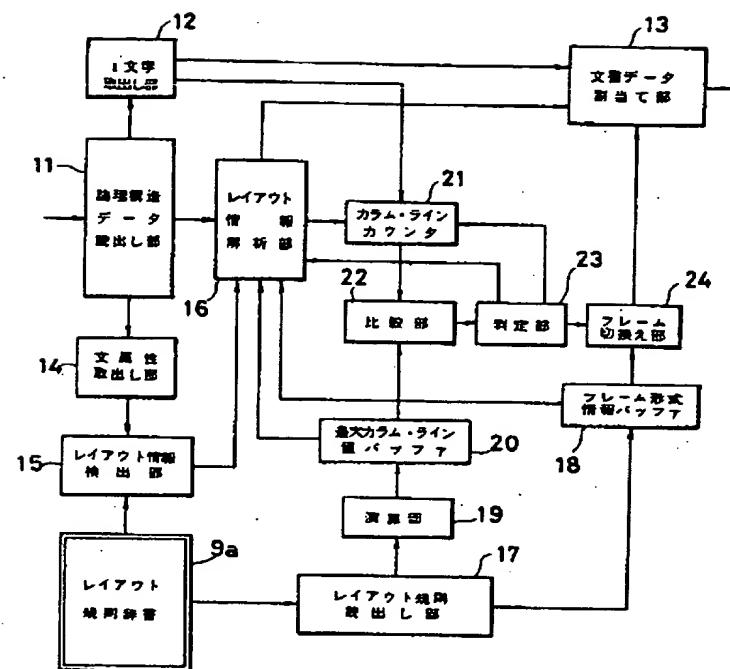
(d)

第6図

【文書検索システム】
 (1.title/z22:1 .z23:10)
 【大川太郎】
 (1.author/z22:1 .z23:11)
 【4.はじめ】
 (1.chapter/z4:1 .z7:7 .z21:1 .z22:3 .z23:6 .z24:2)
 【この発明は.....】
 (1.paragraph/z28:1 .z29:)
 }
 【(2)表示の段階】
 (2.item/z5:1 .z22:2 .z23:6)
 【この種の分野では、技術一般的に.....】
 (2.paragraph/z28:1 .z29:)
 }

第8図

特開昭 62-229364 (12)



第 9 図

(a)

フレーム番号	出力位置	X幅	Y幅	属性
1	X1, Y1	WX1	WY1	Title/author
2	X2, Y2	WX2	WY2	Text/paragraph
3	X3, Y3	WX3	WY3	Text/paragraph

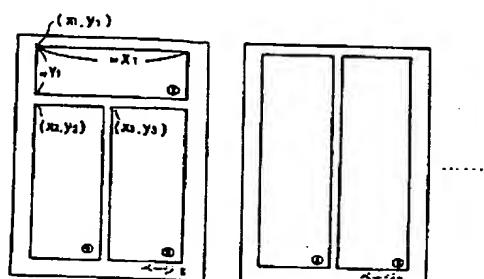
(b)

```

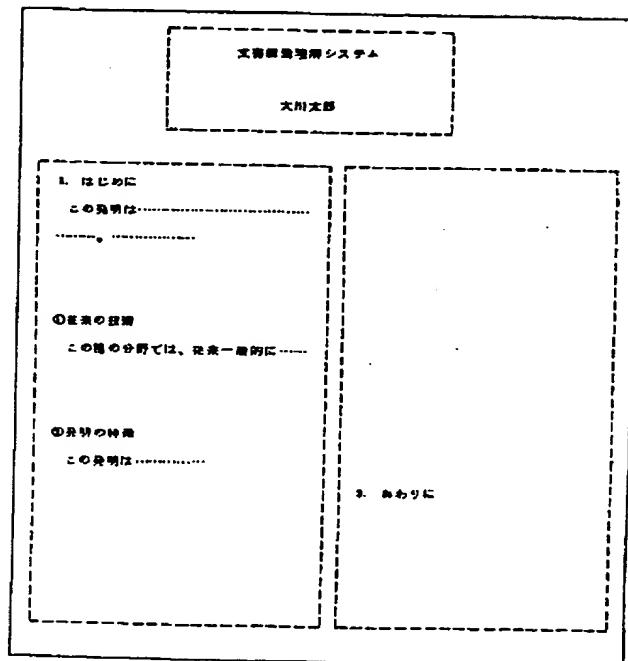
title : #f, #e, ds:1
author : #f, #e, us:1, ds:1
index : #f, #s:1, us:1, ds:1
pindex : #f, #s:3, us:1
paragraph : #f, #s:1
item : #f, #s:4
paragraph : #f, #s:4

```

第 10 図



第 11 図



第 12 図